

REUTERS/Paolo Whitaker

Data Citation Index

Todd King

PDS/PPI

UCLA

Megan Force

Digital Research Analyst - Physical Science

Thomson Reuters

ESSENCE OF THE DATA CITATION INDEX

≈

Data are raised to a first class
research object alongside the
literature

≈

Encourage formal citations
for data objects

≈

DATA CITATION INDEX AIMS

- ▶ Enable the discovery of data repositories, data studies and data sets in the context of traditional literature
- ▶ Link data to research publications
- ▶ Help researchers find data sets and studies and track the full impact of their research output
- ▶ Provide expanded measurement of researcher and institutional research output and assessment
- ▶ Facilitate more accurate and comprehensive bibliometric analyses



Launched October 2012

4 M data records

DRIVEN BY HIGH LEVEL POLICY

The Open Government Partnership:

Second Open Government National Action Plan for the United States of America. http://www.whitehouse.gov/sites/default/files/docs/us_national_action_plan_6p.pdf. Accessed 16 March 2014

NSF Data Management Plan Requirements.

<https://www.nsf.gov/eng/general/dmp.jsp>. Accessed 16 March 2014

Public Library of Science (PLoS) data policy

Bloom T, Ganley E, Winker M (2014) Data access for the open access literature: PLoS'S data policy. PLoS Biol 12(2):e1001797. doi:10.1371/journal.pbio.1001797

Canada's Action Plan in Open Government.

<http://data.gc.ca/eng/canadas-action-plan-open-government>. Accessed 16 March 2014

AGU Data policy (Draft)

<http://publications.agu.org/author-resource-center/publication-policies/data-policy/>



REPOSITORY SELECTION & EVALUATION



Criteria:

- **Material is desirable** to the research community.
- **Persistence and stability** of the repository, with a steady flow of new information.
- **Thoroughness and detail** of descriptive information.
- **Links** from data to research literature.

PDS meets all these →

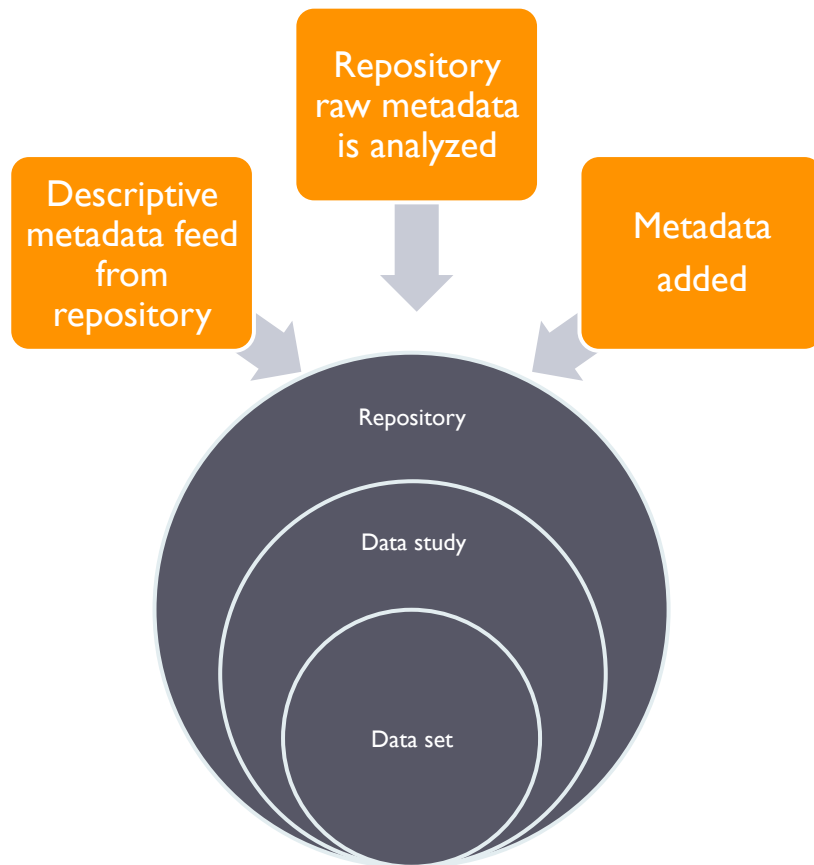
DATA REPOSITORIES

Over 1100 repositories identified



INDEXING A DATA REPOSITORY

ON WEB OF SCIENCE



Record Types

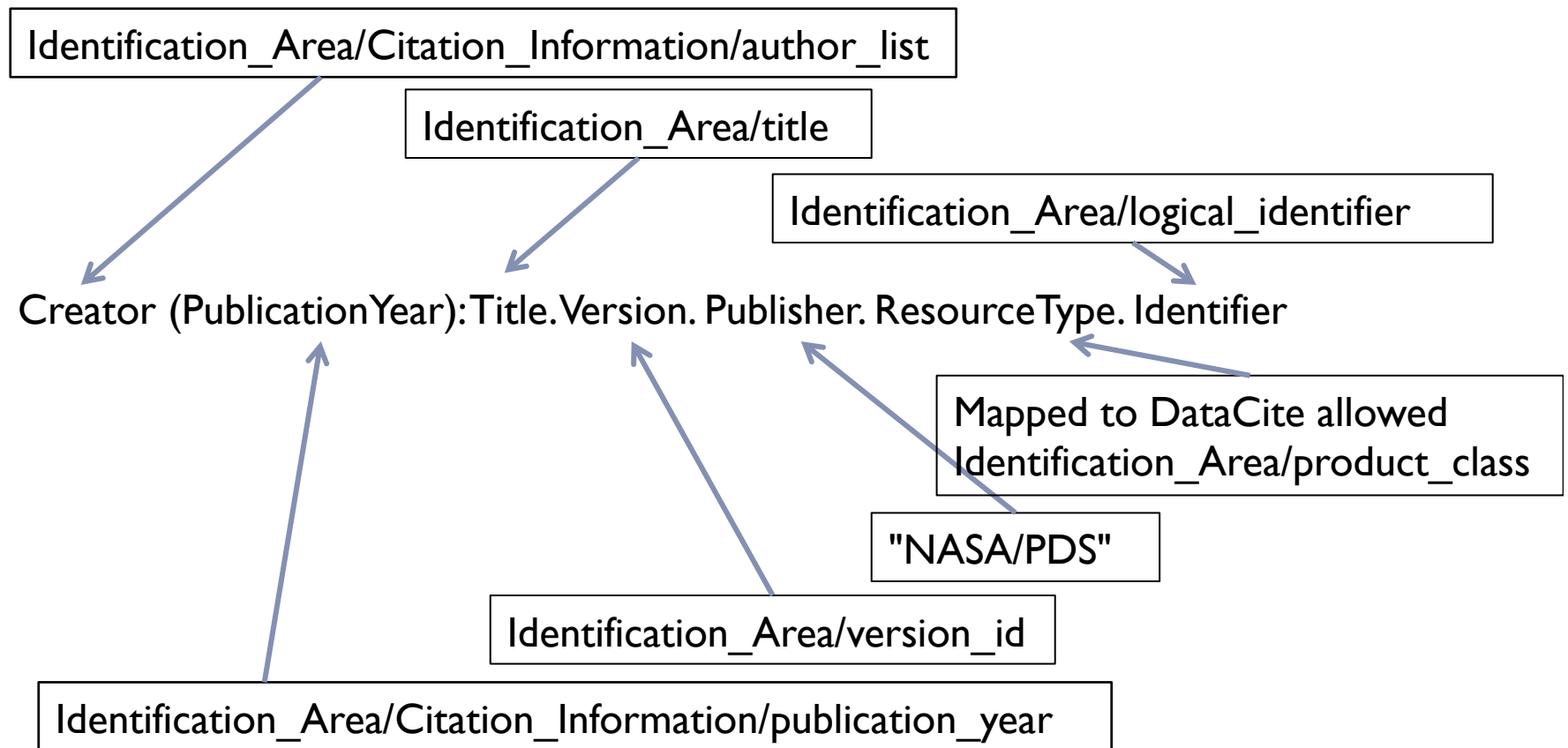
- ▶ **Repository/Source:** Comprises data studies and/or data sets. Stores and provides access to the raw data.
- ▶ **Data Study:** Descriptions of studies or experiments with associated data which have been used in the data study. Includes serial or longitudinal studies over time.
- ▶ **Data Set:** A single or coherent set of data or a data file provided by the repository, as part of a collection, data study or experiment.

REQUIRED METADATA

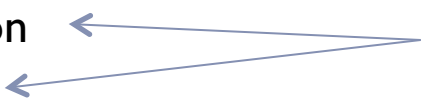
- ▶ Unique ID in repository
- ▶ Date provided
- ▶ Author
- ▶ Repository
- ▶ URL/DOI
- ▶ Title
- ▶ Year Published
- ▶ **Allows creation of a data citation using DataCite guidelines**
Creator (PublicationYear):Title.Version. Publisher. ResourceType. Identifier
- ▶ **Compliance with DataCite Metadata schema v3**
- ▶ **Allows matching of data citations encountered to known data records**

PDS METADATA ALIGNMENT

- ▶ Provide DataCite style citations
- ▶ From Product_Collection:



DATAcite ALLOWED TYPES

- ▶ Audiovisual
 - ▶ Collection
 - ▶ Dataset
 - ▶ Event
 - ▶ Image
 - ▶ InteractiveResource
 - ▶ Model
 - ▶ PhysicalObject
 - ▶ Service
 - ▶ Software
 - ▶ Sound
 - ▶ Text
 - ▶ Workflow
 - ▶ Other
- Matches PDS concepts
- 
- ```
graph LR; A[Matches PDS concepts] --> B[Collection]; A --> C[Dataset];
```

# METRICS AND INSIGHTS

---

Thomson Reuter provides:

- ▶ Times cited (on data page)
- ▶ Citation alerts
- ▶ If the repository agrees to a continued metadata provision for a set period of time (suggested amount is two years), then during that time they will be provided with two logins to DCI, in order to monitor the records, etc.

## Next Steps

---

1. **M/C agrees to collaborate with Thomson Reuters**
  1. **Designate a contact (Andy, Dan, Emily, Sean?).**

### Then...

1. **Provide a metadata sample to Thomson Reuters**
2. **Determine an agreed-upon plan for crosswalking for PDS metadata (collection level)**
3. **Thomson Reuters creates the repository record,**
4. **PDS Provides periodic updates (suggest a 6-month period).**

Thomson Reuters does most of the work!

## PDS MIGHT CONSIDER

---

- I. Require (Strongly Recommend) use of standard index terms (keywords) from community support lists (Thompson Reuters dictionary/thesaurus).



# BACKUP SLIDES

---

# REFERENCES

---

- ▶ **DataCite Metadata Schema v 3.1**  
<http://schema.datacite.org/meta/kernel-3.1/index.html>
- ▶ **Data Citation Index.**  
<http://thomsonreuters.com/data-citation-index/>
- ▶ **Web of Science**, Master Data Repository List:  
[http://wokinfo.com/products\\_tools/multidisciplinary/dci/repositories/search/](http://wokinfo.com/products_tools/multidisciplinary/dci/repositories/search/)
- ▶ Megan M. Force, Nigel J. Robinson (2014): **Encouraging data citation and discovery with the Data Citation Index.** Springer.  
<http://link.springer.com/article/10.1007/s10822-014-9768-5>

# Next Steps

---

1. **Provide a metadata sample to Thomson Reuters**
  - ▶ XML is helpful, but we will accept other forms of tagged metadata, preferably hierarchical.
  - ▶ Enable harvest/delivery through OAI-PMH, ftp, APIs, or by email if the corpus is not overlarge.
2. After analysis of the metadata Thomson Reuters consults with designated individual(s) at the organization to **determine an agreed-upon plan for crosswalking the metadata** into our records.
3. **Thomson Reuters creates the repository record**, ingest the metadata into our system, map the fields, add subject categories, and the records proceed through the pipeline to the live resource.
4. **Provide periodic updates** (suggest a 6-month period).
  - ▶ Thomson Reuters will work with the repository to make this process as automated as possible, in order to minimize effort on the side of the repository.

# REPOSITORY EVALUATION

---

## Data deposit

- Repository must hold data
- Repository must provide access to data

## Active

- Material added/updated
- Provide statistics on deposited data
- Actively curate data in the archive

## Persistent

- Persistent IDs, DOIs or other permanent ID
- Contacts available for confirmation of interpretation
- Indication of intention to preserve data or provide access over the long term
  - Contingency if repository was to cease to operate
- Make data accessible (or state licensing terms)
- Sustainable
  - Funding information available for repository and deposited data

## Proof of reuse

- Links to literature
- Citation in literature databases

# DATA CITATIONS

---

Citations follow the DataCite recommendations:

Creator (PublicationYear):Title. Publisher. Identifier

with optional Version and Resource Type

Creator (PublicationYear):Title.Version. Publisher. ResourceType.  
Identifier

Example:

Irino,T;Tada, R (2009): Chemical and mineral compositions of  
sediments from ODP Site 127-797. Geological Institute,  
University of Tokyo. <http://dx.doi.org/10.1594/PANGAEA.726855>